

Confidence in Cognition and Intrapersonal Perception: Do We Know What We Think We Know About Our Own Cognitive Performance and Personality Traits?

Sandra F. Baker (bakers@usq.edu.au)

Department of Psychology
University of Southern Queensland, Toowoomba QLD 4350 Australia

Gerard J. Fogarty (fogarty@usq.edu.au)

Department of Psychology
University of Southern Queensland, Toowoomba QLD 4350 Australia

Abstract

Calibration research is concerned with the accuracy of confidence judgments made by individuals when responding to various cognitive tasks. Within the cognitive domain, research has demonstrated the existence of a trait of self-confidence that appears to be independent of the type of cognitive activity being investigated. However, the generality of this trait across other domains, such as personality assessment, remains largely unexplored. The present study addressed this by including a number of cognitive and personality assessment tasks within a single battery. It was expected that the usual general self-confidence factor would emerge in the structural analysis of the cognitive tasks and that this factor would also share variance with confidence measures obtained from the personality tasks. This study also investigated whether confidence and calibration differed as a function of ability level. A total of 127 participants completed the battery. Findings indicate that self-confidence did not differentiate from accuracy scores within the cognitive domain and that there was differentiation across the cognitive and personality domains. Also, low scorers were more miscalibrated than high scorers on one of the reasoning tasks.

Introduction

"I think therefore I am..... The end of study should be to direct the mind towards the enunciation of sound and correct judgments on all matters that come before it". Descartes 1596-1650

Sound and correct judgments are just as important today as they were in Descartes' time as we are constantly faced with having to make judgments about some aspect of our daily lives. Whilst Descartes provided a general rule for cognitive activity, the focus of this paper is directed toward judgments that we make about ourselves in terms of our cognitive abilities and personality traits. Accurate self-assessment or self-monitoring of our cognitive performance is a fundamental aspect of successful learning (Flavell, 1977), and may be critical in many work situations. Similarly, research into the accuracy of self-perception has been "an issue of longstanding concern to philosophers and social scientists" (John & Robins, 1994). The calibration paradigm (see below) will be

used to assess self-monitoring within both the cognitive and personality domains.

There are reliable findings that individuals tend to be overconfident or underconfident when evaluating the accuracy of their cognitive performance. Individuals are regarded as mis-calibrated if there is a mismatch between accuracy and confidence. Self-monitoring is operationalised by self-confidence scores, that is, individuals are asked to express how confident they are in the accuracy of their judgments, or answers to every item of a cognitive test. Bias scores indicate the size and direction of this mismatch and is simply the average of confidence ratings across all items in a task minus the proportion correct score for that task (Stankov & Crawford, 1996). A positive bias score suggests overconfidence whereas a negative score indicates underconfidence. Bias scores will be generated for the cognitive variables.

Winman, Juslin, and Bjorkman (1998) found that individuals were well calibrated on tasks that required hindsight bias, and Stankov and Crawford, (1996) found good calibration on the Raven's Progressive Matrices Test and digit span tasks. Individuals are underconfident when answering questions about future events (Vreugdenhil & Koele, 1988), judging line length (Bjorkman, Juslin, & Winman, 1993), however, underconfidence did not generalise to perceptual tasks in other sensory modalities (i.e., auditory, kinesthetic, gustatory, and olfactory) or to other visual perceptual tasks (e.g., a square gap task and the Muller Lyer Illusion) (Stankov, in press). In terms of overconfidence, there is a large body of evidence to suggest that individuals anticipate performance levels higher than their actual results (see review by Kleitman & Stankov, 2001).

Of interest to the present study is that differential psychologists have demonstrated the existence of a trait of self-confidence that appears independent of the type of cognitive activity being investigated (c.f., Stankov, in press; Stankov & Dolph, 2000). Moreover, it has been argued that this general self-monitoring/self-confidence trait represents one aspect of metacognition related to the accuracy of self-assessment in the cognitive domain. Interestingly, the generality of this trait sets it apart from other known constructs such as self-efficacy where confidence is domain dependent. However,

differential psychologists have so far failed to investigate the generality of this trait in other domains. Stankov (1999) has placed the confidence trait somewhere between the boundaries of intelligence and personality. If this is the case, it is possible that such a general trait could cut across into the personality domain if one refers to global self-concept theories where a collection of beliefs about oneself includes perceptions of appearance, personality, and abilities (Shavelson, Hubner, & Stanton, 1976). Investigating the generality of the self-confidence trait in the personality domain has important implications for calibration theorists trying to understand self-monitoring as current theoretical propositions cannot account for a number of research findings.

Research Aims

In order to assess confidence in the personality domain, five personality descriptions were constructed based on the trait adjectives from the International Personality Item Pool Five-Factor Personality Scale (IPIP; Goldberg, 1997) that assesses the dimensions of Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Intellect. The IPIP adjectives were chosen because the personality dimensions are positive thereby lessening self-enhancement biases. A confidence measure will be obtained after participants indicate whether each personality description is an accurate description of that aspect of their personality. It was expected that the usual general self-confidence factor would emerge in the structural analysis of the cognitive tasks and that this factor would also share variance with confidence measures obtained from the personality tasks. One cognitive bias factor is also expected.

Other individual differences apparent in the confidence literature of interest to the present study are that males tend to be more confident than females on cognitive tasks (Pulford & Colman, 1997) and that older individuals are more confident than younger individuals with cognitive tasks (Crawford & Stankov, 1996). We are also interested in the possibility that individual differences in confidence and mis-calibration could be linked to ability levels. Spence (1996) reported that experts are better calibrated than novices but the very early empirical literature on confidence suggests that those of lower ability tend to be more confident, a finding that is counterintuitive. More recently, Kruger and Dunning (1999) investigated whether low scorers on tests of humour, grammar and logic provided accurate self-assessments of their abilities. Results indicate that low scoring individuals "grossly overestimated their test performance and ability" (p.1121). Consequently, this important issue will be addressed in the current study by investigating whether significant differences are apparent in mis-calibration amongst participants who scored in the bottom quartile on a general knowledge test versus those who scored in the top quartile. It is predicted that low scorers will be more mis-calibrated than high scorers on tasks of cognitive abilities. A Multivariate Analysis of Variance will be used in order to compare mean bias scores. In a similar vein to Kruger and Dunning, tasks of

acculturated knowledge and reasoning will be included in the test battery. Correlations will be used to examine the relationship between ability and self-confidence.

Method

Participants

A total of 127 individuals participated in this study. The sample comprised 40 males and 87 females, ranging in age from 17 to 74 years ($M = 34.42$, $SD = 12.76$). The highest educational level of the sample varied from completion of grade 7,8,9,10,11, or 12 ($n = 77$), to completion of tertiary studies ($n = 50$). Sixty participants were enrolled in undergraduate Psychology courses at the University of Southern Queensland and received course credit in return for their participation. Snow ball sampling techniques were used to obtain a community sample of 67 participants. The experimenter randomly approached community organisations and provided individuals with a rationale for the study and asked respondents if they could recommend others who might be interested in participating in this research. The community sample came from both metropolitan and regional areas in Queensland.

Materials

Demographic questions consisted of items regarding, gender, age, and highest level of education. All participants completed a battery of five cognitive tasks, and five short descriptions of personality. Each of these measures is described below.

General Knowledge Test (GKT-Stankov, 1997). This test covers knowledge of diverse areas such as history, and geography.

Letter Series Test (LST-Stankov, 1997). Participants were presented with a series of letters (e.g., A, D, G, J, ?) and were asked to provide the next letter of the series. A time limit of four minutes was given.

Concealed Words Test (CWT-Stankov, 1997). Participants were asked to identify words when parts of each letter were degraded. Participants were given a time limit of two minutes.

Esoteric Analogies Test (EST-Stankov, 1997). Participants were asked to choose words that completed verbal analogies. Four response options were provided. For example, LIGHT is to DARK as HAPPY is to GLAD, SAD, GAY, EAGER. A four minute time limit was imposed.

Cattell's Matrices (CM-Stankov, 1997). Participants were asked to choose from among six options the design that completes a matrix. Four minutes were allowed to complete this task.

Participants were asked to provide an answer to every trial of the aforementioned tasks as well as a confidence rating indicating how confident they were that the answer provided was correct. For the open ended tests (i.e., GKT & LST) confidence ranged from 0% (Just guessing) to 100% (Absolutely certain). For the other multiple choice tests, the starting point on the confidence scale was 100/k, where k = the number of response alternatives.

Self-Rated Personality Descriptions based on the Trait Adjectives by Goldberg (1997). Five personality descriptions were constructed from the 100 item IPIP scale (Goldberg, 1997). For each of the five personality dimensions participants were presented with a personality description and were asked to rate the extent to which the overall description generally reflected their personality on a 11-point scale from -5 (Not like me) to 5 (Like me). For example, the Extraversion personality description appears below:

"I don't mind being the centre of attention; I make friends easily; I take charge; I know how to captivate people; I feel at ease with people; I am skilled in handling social situations; I am the life of the party; I start conversations."

Following each description, participants provided their confidence rating after reading the following:

"Imagine that there was some device that could accurately tell us about your personality. How confident are you that the rating you gave above would correspond with the device's rating? Please rate your confidence on the scale that appears below by circling your level of confidence."

The confidence scale for the personality judgments, used 10% intervals and ranged from 0% (Just guessing) to 100% which indicated absolute certainty.

Procedure

Participants were tested individually after providing a rationale and explaining that all data would remain confidential. The test battery started with the GKT test for all participants. The order of the qualitative descriptors was randomised for each participant with the only constraint being that each descriptor be followed by a cognitive task. Participants were not informed of their accuracy during testing but were given feedback at the end of the experiment. The battery took approximately one hour to complete.

Results

Reliability coefficients, descriptive statistics, and correlations among accuracy, confidence and personality variables are presented in Table 1. All reliability estimates fall within acceptable limits for experimental research (Gregory, 1996). The factorial structure of the test battery was investigated using principal components analysis with oblimin rotation using the cognitive accuracy, cognitive confidence, and personality confidence variables. The CWT accuracy and confidence scores were deleted from the analysis as their measures of sampling adequacy fell below the recommended cut-off of .5 (Tabachnick & Fidell, 1996). Three factors emerged, accounting for 67% of the total variance. The first factor was labelled acculturated knowledge as both the accuracy and confidence scores from the GKT and EST tasks loaded highly on it. The second factor had high loadings from all of the confidence ratings from the personality descriptors and was labelled personality confidence. A reasoning factor emerged as the last factor with high loadings from both the LST and CM accuracy and confidence scores. As

mathematical dependence exists between the bias, accuracy and confidence scores a separate principal components analysis was undertaken using only the cognitive bias scores. One cognitive bias factor emerged from the data set accounting for 45.6% of the total variance.

The intercorrelations between ability (measured by accuracy) and confidence scores in Table 1 are of interest to this study. Strong correlations (i.e., $r_s > .57$) between accuracy and confidence scores were apparent for each of the tasks except for the CM test ($r = .38$, $p = .01$). This indicates that for most tasks those with higher ability were also more confident in their performance on those tasks.

Participants' scores on the GKT task were divided into quartiles in order to investigate whether low scorers were more mis-calibrated than high scorers. A Multivariate Analysis of Variance was performed on the four dependent variables: LST, CWT, EST and CM bias scores with the independent variable being GKT quartile (first or fourth). Pillai's F statistic was chosen as it is the most robust against violations of homogeneity of variance-covariance matrices (Tabachnick & Fidell, 1996). Using Pillai's Trace, a difference between low and high scorers was found on the combined DVs $F(4, 64) = 9.99$, $p < .05$, with partial $\eta^2 = .15$. Thus, 15% of the variance on the composite Dvs can be uniquely attributed to differences between low and high scorers. As there was no theoretical justification for ordering the Dvs a bonferroni adjustment ($\alpha = .01$) was used to evaluate univariate analyses. This resulted in one significant difference between the means for EST bias scores ($F(1, 67) = 8.802$, $p < .004$, with partial $\eta^2 = .12$). This indicates that low scorers were significantly more biased on the EST than were high scorers and that 12% of the variance can be attributed to the differences between these groups.

Discussion

The present study was designed to investigate whether the cognitive trait of self-confidence generalised to the domain of personality assessment, and whether individual differences in ability levels contributed to mis-calibration and to confidence ratings. Contrary to expectations, a general confidence factor did not emerge from this data set. Confidence and accuracy scores loaded together and formed two separate factors (i.e., acculturated knowledge and reasoning). Whilst this finding is inconsistent with some of the research literature (see Stankov, in press for a review), it is not unheard of as there are instances in the literature where a general confidence factor failed to emerge when correlations between confidence and accuracy scores were greater than .5 (c.f., Stankov & Dolph, 2000). In the present study, the average correlation between confidence and accuracy scores was .66 thereby providing one explanation for our inconsistent findings. In terms of the bias scores, however, our results are consistent with previous research findings where bias scores from differing cognitive domains loaded together

Table 1: Descriptive Statistics, Reliability Coefficients, and Correlations between Main Variables (N = 127)

	<i>M</i>	<i>SD</i>	α	1	2	3	4	5	6	7	8	9	10
Cognitive Accuracy													
1. GKTAC (20 items)	.40	.20	.79	1.00									
2. LSTAC (15 items)	.61	.18	.66	.20	1.00								
3. CWTAC (26 items)	.26	.11	.64	.19	.13	1.00							
4. ESTAC (24 items)	.60	.16	.73	.61	.27	.13	1.00						
5. CMAC (11 items)	.76	.15	.55	.32	.44	.15	.23	1.00					
Cognitive Confidence													
6. GKTC	.45	.20	.86	.84	.19	.16	.60	.25	1.00				
7. LSTC	.61	.15	.78	.25	.58	.17	.34	.36	.33	1.00			
8. CWTC	.26	.12	.78	.04	-.03	.85	.09	-.01	.12	.15	1.00		
9. ESTC	.66	.17	.90	.41	.09	.07	.69	.07	.53	.33	.19	1.00	
10. CMC	.86	.13	.86	.31	.24	.19	.30	.38	.37	.54	.23	.52	1.00
Personality Confidence													
11. CONC	-	-	-	-.20	-.03	-.15	-.05	-.23	-.19	-.07	-.12	.06	.01
12. EMOTC	-	-	-	-.02	.09	.07	.09	-.06	-.03	.02	.11	.10	.10
13. INTELLC	-	-	-	.04	.01	.09	.12	.04	.07	.01	.11	.15	.22
14. EXTRAC	-	-	-	-.17	.09	-.04	.04	.01	-.11	-.06	-.04	.10	.09
15. AGREEC	-	-	-	-.12	-.04	-.08	.02	-.17	-.09	.04	.00	.12	.09
Table 1 (contd)	<i>M</i>	<i>SD</i>	α	11	12	13	14	15					
11. CONC				1.00									
12. EMOTC				.49	1.00								
13. INTELLC				.38	.42	1.00							
14. EXTRAC				.46	.41	.32	1.00						
15. AGREEC				.49	.55	.54	.36	1.00					

Note. $r_s \geq .16$ are significant at the .05 level, and $r_s \geq .23$ are significant at the .01 level. GKT= General Knowledge; LST = Letter Series; EST = Esoteric Analogies; CWT = Concealed Words; CM = Cattell's Matrices; CONC= Conscientiousness; EMOTC= Emotional stability; INTELLC = Intellect; EXTRAC = Extraversion; AGREEC= Agreeableness.

onto one factor (c.f., Stankov, in press; Stankov & Dolph, 2000). Our results indicate that if individuals were mis-calibrated in the reasoning domain then they were also mis-calibrated in the acculturated knowledge domain.

Regarding personality confidence, a separate factor emerged that was essentially uncorrelated ($r_s < .002$) with the cognitive factors. This finding does not fit with a model that shows a confidence trait sitting between the personality and ability domains (Stankov, 1999). However, there are some methodological issues to be resolved regarding the assessment of confidence in the personality domain and it is still unclear whether personality confidence shares variance with a confidence factor derived from the ability domain. Future research could address this issue by including a diverse battery of cognitive and personality assessment tasks and by trialing different methods of assessing confidence in personality judgments. Interestingly, correlations between confidence ratings in the personality domain were as high if not higher than correlations between confidence ratings from the cognitive tasks suggesting the presence of a general confidence factor in personality judgments. Indeed, we have replicated these correlations in a subsequent study. However, the generality of a personality confidence

factor across other personality judgments remains unexplored. Perhaps it could be argued by some calibration researchers that the personality confidence judgments were of a global nature and may be more similar to post-test evaluative confidence judgments in the cognitive domain. With this type of judgment individuals are asked to predict how many items they believed were correct after completing a test and are then asked to provide a confidence rating. Factor analyses of post-test confidence and item-by-item confidence estimates in the cognitive domain results in two separate factors. We therefore examined in a subsequent study whether there is differentiation between item-by-item confidence estimates versus global confidence estimates in the personality domain and preliminary analyses indicate one confidence factor.

There was partial support for the hypothesis that low scorers on tasks of cognitive abilities would be more mis-calibrated than high scorers. Significant differences were found for the EST task that taps both reasoning acculturated knowledge. Low scorers were overconfident (bias score = 11) whereas high scorers were well calibrated. Perhaps this task was novel to a number of the low scorers in the community sample and therefore they were unaware of how they

performed. The finding for the EST task parallels the work of Kruger and Dunning (1999) who found that low scorers overrated their test performance. For the other tasks, there were no significant differences in terms of mis-calibration between low and high scorers. Perhaps this could be attributed to sample differences as the present study comprised both university students and individuals from the general population whereas Kruger and Dunning's sample were university students. In terms of the relationship between ability and self-confidence, we found that those with higher ability were also more confident in their performance. This finding is consistent with expectations but modifies claims by researchers such as Kruger and Dunning that less capable people are more confident than they should be about their performance. We have found some evidence of this tendency through the analyses of bias scores but these correlations between confidence and accuracy show that those who know more are also more confident about that knowledge.

So, do we know what we think we know about our cognitive performance? Bias scores indicated good calibration for all tasks except the CM test. A bias score of 10 for the CM test is consistent with previous research using tests of a similar nature (e.g., Pallier et al., 2002) and in this sample participants appeared unaware that they had made errors. With regard to personality, the answer remains unclear as we do not yet have an objective criterion by which to determine the accuracy of personality judgments. We are currently working on other methods of deriving these measures.

References

- Bjorkman, M., Juslin, P., & Winman, A. (1993). Realism of confidence in sensory discrimination: The underconfidence phenomenon. *Perception and Psychophysics*, 54, 75-81.
- Crawford, J. D., & Stankov, L. (1996). Age Differences in the Realism of Confidence Judgements: A Calibration Study using Tests of Fluid and Crystallized Intelligence. *Learning and Individual Differences*, 8, 83-103.
- Flavell, J. H. (1977). *Cognitive development*. Englewood Cliffs, NJ: Prentice-Hall.
- Goldberg, L. (Ed.). (1997). *A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models*. The Netherlands: Tilburg University Press.
- Gregory, R. J. (1996). *Psychological testing: History, principles, and applications*. Boston: Allyn & Bacon.
- John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology*, 66, 206-219.
- Kleitman, S., & Stankov, L. (2001). Ecological and person-oriented aspects of metacognitive processes in test-taking. *Applied Cognitive Psychology*, 15 (3), 321-341.
- Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134.
- Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, S., Stankov, L., & Roberts, R. (2000). The role of individual differences in the realism of confidence judgements. *The Journal of General Psychology*, 129(3), 257-299.
- Pulford, B. D., & Colman, A. M. (1997). Overconfidence: Feedback and Item Difficulty Effects. *Personality and Individual Differences*, 23, 125-133.
- Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46, 407-441.
- Spence, M. T. (1996). Problem-problem solver characteristics affecting the calibration of judgments. *Organizational Behavior and Human Decision Processes*, 67 (3), 271-279.
- Stankov, L. (1997). The Gf/Gc Quickie Test Battery: Sydney, Australia: The University of Sydney, School of Psychology. Unpublished test battery.
- Stankov, L. (1999). Mining on the "no man's land" between intelligence and personality. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and Individual Differences: Process, trait, and content determinants* (pp. 314-337). Washington, DC: American Psychological Association.
- Stankov, L. (in press). Accuracy, confidence, and speed in perceptual tasks: Calibration effects and individual differences. *Journal of Experimental Psychology: Human Perception and Performance*. In press.
- Stankov, L., & Crawford, J. (1996). Confidence judgements in studies of individual differences support the 'Confidence/frequency effect'. In C. Latimer & J. Mitchell (Eds.), *At Once Scientific and Philosophic: A Festschrift in Honour of J. P. Sutcliffe*. (pp. 215-213). Sydney: University of Sydney Press.
- Stankov, L., & Dolph, B. (2000). Metacognitive aspects of test-taking and intelligence. *Psychologische Beitrage, Band 42*, 213-227.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). Sydney: Harper Collins.
- Vreugdenhil, H., & Koele, P. (1988). Underconfidence in predicting future events. *Bulletin of the Psychonomic Society*, 26(3), 236-237.
- Winman, A., Juslin, P., & Bjorkman, M. (1998). The Confidence-Hindsight Mirror Effect in Judgement: An Accuracy-Assessment Model for the Knew-It-All-Along Phenomenon. *Journal of Experimental Psychology*, 24(2), 415-431.